



Savage, J. T. S., Bates, P., Freer, J., Neal, J., & Aronica, G. T. (2016). When does spatial resolution become spurious in probabilistic flood inundation predictions? *Journal of Hydrology*, 30(13), 2014-2032. DOI: 10.1002/hyp.10749

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1002/hyp.10749](https://doi.org/10.1002/hyp.10749)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <http://onlinelibrary.wiley.com/wol1/doi/10.1002/hyp.10749/abstract>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/pure/about/ebr-terms.html>

When does spatial resolution become spurious in probabilistic flood inundation predictions?

James Thomas Steven Savage,^{1*} Paul Bates,¹ Jim Freer,¹ Jeffrey Neal¹ and Giuseppe Aronica²

¹ *Geographical Sciences, University of Bristol, Bristol, United Kingdom*

² *Department of Civil Engineering, University of Messina, Italy*

Abstract:

Advances in remote sensing have enabled hydraulic models to run at fine scale resolutions, producing precise flood inundation predictions. However, running models at finer resolutions increase their computational expense, reducing the feasibility of running the multiple model realizations required to undertake uncertainty analysis. Furthermore, it is possible that precision gained by running fine scale models is smoothed out when treating models probabilistically. The aim of this paper is to determine the level of spatial complexity that is required when making probabilistic flood inundation predictions. The Imera basin, Sicily is used as a case study to assess how changing the spatial resolution of the hydraulic model LISFLOOD-FP impacts on the skill of conditional probabilistic flood inundation maps given model parameter and boundary condition uncertainties. We find that model performance deteriorates at resolutions coarser than 50 m. This is predominantly caused by changes in flow pathways at coarser resolutions which lead to non-stationarity in the optimum model parameters at different spatial resolutions. However, although it is still possible to produce probabilistic flood maps that contain a coherent outline of the flood extent at coarser resolutions, the reliability of these maps deteriorates at resolutions coarser than 100 m. Additionally, although the rejection of non-behavioural models reduces the uncertainty in probabilistic flood maps the reliability of these maps is also reduced. Models with resolutions finer than 50 m offer little gain in performance yet are more than an order of magnitude computationally expensive which can become infeasible when undertaking probabilistic analysis. Furthermore, we show that using deterministic, high-resolution flood maps can lead to a spurious precision that would be misleading and not representative of the overall uncertainties that are inherent in making inundation predictions. Copyright © 2015 The Authors Hydrological Processes Published by John Wiley & Sons Ltd.

KEY WORDS hydraulic modelling; uncertainty; flood inundation; spatial resolution; probabilistic; epistemic errors

Received 13 May 2015; Accepted 10 November 2015

INTRODUCTION

There have been a number of high profile flood events over the last decade throughout the world that have demonstrated the large damage to the economy and humanity that flooding can cause (Munich RE, 2014). Climate change has already been shown to have likely contributed to extreme flood events (Pall *et al.*, 2011) and as climate continues to warm (IPCC, 2013) extreme events are predicted to increase in their frequency and severity. Consequently research is required to improve predictions of flood hazard in order to help mitigate against significant flood events in the future. Over the last decade, the increased availability of high-resolution topographic and validation data collected through remote sensing and field survey (Bates, 2004, 2012) has helped to improve hydraulic models and allow them to be scrutinized and validated more rigorously. However,

subsequent flood hazard predictions remain subject to large amounts of uncertainty, and it is vital that we improve our knowledge of the impact uncertainties have in order to inform more effective decision making both in the present and future.

Flood inundation modelling

There are many methods available to assess flood hazard; the most common being flood frequency analysis (Blazkova and Beven, 2009a; Merz and Thielen, 2009) and flood inundation modelling (for example Bates and De Roo, 2000; Mignot *et al.*, 2006; Yu and Lane, 2006; Fewtrell *et al.*, 2008; Neal *et al.*, 2011). Hydraulic models allow spatial predictions of flood inundation to be made and studies have applied such models in both rural and urban environments and for a variety of floods, including fluvial, coastal and pluvial (Bates and De Roo, 2000; Horritt and Bates, 2001a; Mignot *et al.*, 2006; Yu and Lane, 2006; Fewtrell *et al.*, 2008; Purvis *et al.*, 2008; Neal *et al.*, 2011; Sampson *et al.*, 2012; Falter *et al.*, 2013; Quinn *et al.*, 2013; Tarpanelli *et al.*, 2013). The primary

*Correspondence to: James Thomas Steven Savage, Geographical Sciences, University of Bristol, Bristol, United Kingdom.
E-mail: James.Savage@bristol.ac.uk

data required to run and validate hydraulic models are topography, boundary conditions and validation data. The topography of a region is typically collected through remote sensing and defined as elevation points in a Digital Elevation Model (DEM). Light Detection and Ranging (LiDAR) technologies have allowed surface elevation data to be measured with a horizontal resolution of less than 0.5 m with very small vertical errors (± 0.15 m). Representing topography at such fine resolutions allows very precise predictions of inundation to be made which has helped to improve the skill of urban scale inundation modelling (Fewtrell *et al.*, 2008; Neal *et al.*, 2011; Aronica *et al.*, 2012). Boundary conditions required by hydraulic models are typically discharge data, tidal data or point sources of flood waters, for example a burst pipe.

A variety of data exist to help quantify how well hydraulic models perform, including discharge data which allow the peak discharge and time to peak discharge to be assessed (Aronica *et al.*, 1998), satellite images of flood extent which are used to delineate the inundation shoreline change over time (e.g. Tarpanelli *et al.*, 2013), variable interferometric radar observations of water elevation (Jung *et al.*, 2012), observations of water level from marks left on structures and wrack marks found along the boundary of the flood extent (Werner *et al.*, 2005; Neal *et al.*, 2009, 2011; Parkes *et al.*, 2013). The use of observed water depths to constrain hydraulic models has been shown to improve the reliability and skill of hydraulic models used for flood risk assessments and aid constraining parametric uncertainty (Werner *et al.*, 2005; Neal *et al.*, 2009, 2011; Stephens *et al.*, 2012), whilst the use of synthetic aperture radar (SAR) imagery even at coarse resolutions has been shown to improve the calibration of hydraulic models (Tarpanelli *et al.*, 2013). The availability of validation data allows competing models to be compared with one another and therefore permits the quantification of the impact of uncertainties on model predictions.

Uncertainty in flood inundation modelling

Despite the increase in data availability, predictions of flood hazard continue to be subject to a number of uncertainties. Improving our understanding of these uncertainties is vital to allow informed flood risk decision making that reflects the inherent uncertainties. These uncertainties can be classified as either aleatory or epistemic. Aleatory uncertainties exist as a result of the randomness of systems and processes (Beven *et al.*, 2011a). One way of estimating them is to assess the probability of event recurrences using flood frequency analysis (Merz and Thielen, 2009). Epistemic uncertainties are attributable to the imperfect knowledge of the processes and systems being modelled (Beven *et al.*,

2011a). Although it can be difficult to quantify these uncertainties, it is important that they are considered as they may contribute the greatest error to model predictions.

Bayesian techniques have been used to try to quantify epistemic uncertainties (Hall *et al.*, 2011); however, these techniques may not always be appropriate because of the structured and non-stationary nature of these uncertainties (Beven *et al.*, 2011b). There are alternative, less formal approaches available to quantify the affect epistemic uncertainties have on predictions. One example is the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992) which is a method that has been commonly utilized in hydrology (Aronica *et al.*, 1998, 2002; Beven and Freer, 2001; Romanowicz and Beven, 2003; Blazkova and Beven, 2009b). Another approach is sensitivity analysis, which has been used to understand the dominant uncertainties in predicting flood hazards (Hall *et al.*, 2005, 2009; Pappenberger *et al.*, 2008).

Some of the epistemic uncertainties that studies have attempted to quantify using the GLUE methodology include boundary condition uncertainty (Di Baldassarre and Montanari, 2009; Domeneghetti *et al.*, 2013), the choice or structure of model used to produce flood hazard outputs (Hunter *et al.*, 2008; Neal *et al.*, 2012), the parameterization of the model (Aronica *et al.*, 2002) and SAR data used for model validation (Di Baldassarre *et al.*, 2009; Stephens *et al.*, 2012).

The variation in flood hazard predictions caused by epistemic uncertainties means that within the bounds of the epistemic uncertainties, there are multiple plausible model realizations. Any predictions should account for this equifinality (Beven and Freer, 2001; Beven, 2006) by considering multiple acceptable models of the system (Aronica *et al.*, 2002) rather than choosing a singular 'best performing' simulation. Furthermore, the non-stationary elements of these uncertainties mean that when future scenarios are applied to a range of acceptable models, their solutions can diverge giving multiple future scenarios which may not have been apparent when analysing the deterministic simulation alone (Cloke *et al.*, 2013; Smith *et al.*, 2014). Consequently, making deterministic high-resolution predictions of flood hazard that do not consider the underlying inherent uncertainties can result in model output that is spuriously precise (Dottori *et al.*, 2013).

Incorporating high-resolution topography when quantifying epistemic uncertainties

The incorporation of high-resolution topography facilitates an improved representation of the wetting and drying processes (Neal *et al.*, 2011) and allows smaller features, for example buildings and kerbs to be explicitly included in the DEM (Fewtrell *et al.*, 2008, 2011a).

However, incorporating high-resolution topography increases the computational expense of model simulations as a result of a higher number of calculations and lower stable time step required to ensure numerical stability. Typically this means that halving the grid resolution increases the run time by an order of magnitude, meaning that running numerous simulations at high resolutions to quantify effective prediction uncertainties may not always be feasible. This becomes a particularly important consideration when multiple simulations are required for a probabilistic analysis. For example calculating flood risk as a result of possible dike breaches (Apel *et al.*, 2009a) or cascading uncertainties for future predictions of flood risk from an ensemble of driving climate data (Smith *et al.*, 2014) requires many model realizations. Additionally it may not always be necessary to use the most complex hydraulic model possible (Apel *et al.*, 2009b). Consequently it is clear that when undertaking probabilistic analysis a trade-off must be made between model complexity and the number of model realizations simulated.

Modellers have explored a number of approaches to reduce computation time. These include running models with a spatially variable timestep (Sanders, 2008), using models with reduced physics (Horritt and Bates, 2001b; Bates *et al.*, 2010; Neal *et al.*, 2012), applying a correction to coarse resolution simulations using a smaller subset of high-resolution model simulations (Néelz *et al.*, 2007), splitting the topography into larger zones around topographic depressions (Gouldby *et al.*, 2008; L'homme *et al.*, 2008; Falter *et al.*, 2013; Jamieson *et al.*, 2012), or in urban regions, using alternative methods of representing urban structures, such as adjusting the friction parameter or using a porosity based approach to replicate buildings as opposed to using high-resolution topography (Dottori and Todini, 2013). A commonly used methodology to reduce simulation time is to run models at coarser spatial resolutions. However previous studies that have analysed how predictions of flood inundations are influenced by the spatial resolution of the model have primarily only looked at the effect on deterministic simulations (Horritt and Bates, 2001a; Yu and Lane, 2006; Fewtrell *et al.*, 2008). Horritt and Bates (2001a) found for a study on the Severn catchment that model performance increased as resolution became finer up to a resolution of 100 m where performance gain plateaued. In this study model performance was assessed against flood extent and floodwave travel time, rather than against water levels which has previously been shown to be a more stringent and robust assessment of model performance (Stephens *et al.*, 2012; Stephens *et al.*, 2014). Yu and Lane (2006) find for an urban environment that although their model is sensitive to spatial resolution, the effect of parameters can offset the loss in performance because of resolution. However

Fewtrell *et al.* (2008) found that for urban areas, the representation of buildings is important and hence finer resolutions are required. They also find that the response of Manning's friction coefficient to model resolution is non-stationary; however, they do not evaluate their model to observed data here.

None of these studies have included an assessment of the influence of observation uncertainty at different spatial resolutions in combination with parametric uncertainty which may lead to interesting changes in interactions as resolution is coarsened. A previous study by Savage *et al.* (2014) that attempted to explore the effect of spatial resolution probabilistically was limited in that only a small number of spatial resolutions were explored where the model performed similarly across each resolution. Therefore as of yet it is not clear how coarse the spatial resolution of a hydraulic model can be, and therefore how much the computational cost of running simulations can be reduced, without substantially degrading the performance of the model given the fuzziness introduced by epistemic uncertainties.

Paper aims

The aim of this study is to evaluate how probabilistic inundation maps are affected by spatial resolution. The hydraulic model LISFLOOD-FP (Bates *et al.*, 2010) will be utilized to simulate the very large flood event that occurred during 1991 in the Imera basin, Sicily. LISFLOOD-FP will be used to produce a series of weighted conditional probability of inundation maps using the GLUE methodology for different grid scale resolutions and a commonly applied series of uncertain boundary conditions. The resultant uncertain predictions will then be compared for each resolution using a performance metric that takes into account uncertainties in observations of water depth that are used to benchmark model performance. This will allow their changing predictive skill to be assessed and the limit at which including more complex topography no longer improves predictive behaviour to be determined. The results of this study will inform flood risk decision makers of the level of topographic complexity necessary to make flood hazard predictions given inherent uncertainties regarding the boundary conditions and parameters.

STUDY SITE—IMERA BASIN, SICILY

The Imera basin in Southern Sicily (Figure 1) is over 2000 km² in size (Aronica *et al.*, 1998) and is a predominantly rural catchment with the main urban developments located along the coast. It is through one of these coastal developments, Licata, where the Imera River, after flowing through the basin, leaves the southern

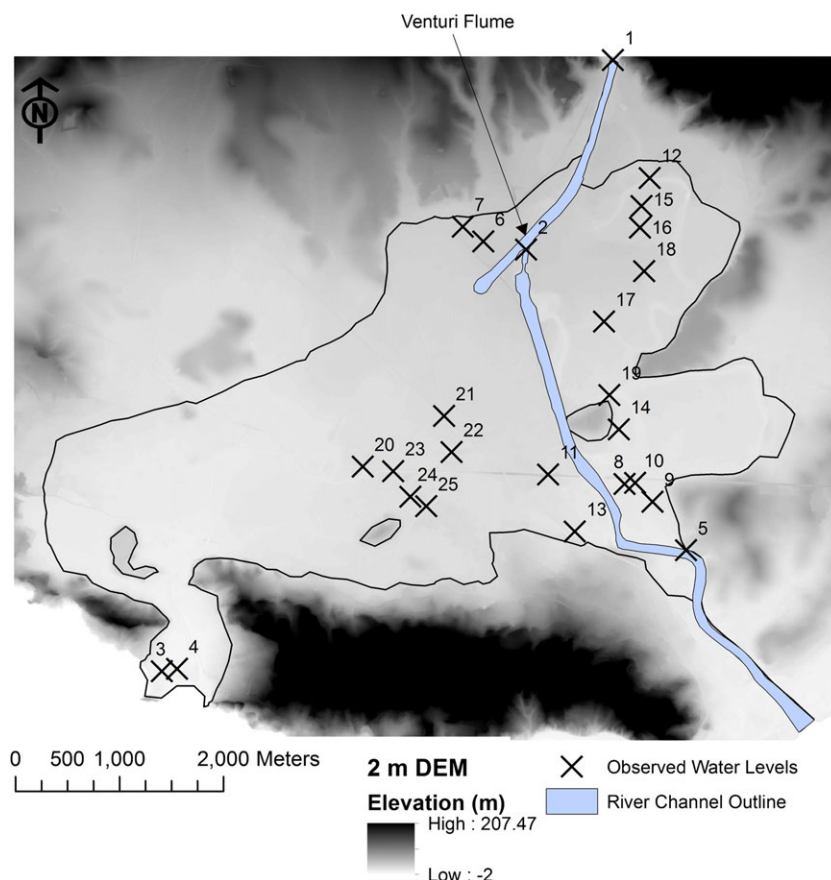


Figure 1. Digital Elevation Model (DEM) of the Imera basin at a 2 m resolution. The crosses indicate locations of water height observations, the solid black line is the observed flood extent and the shaded blue region is an outline of the Imera river

coast and enters the Mediterranean Sea. Towards the south of the basin just a few kilometres upstream of the river mouth there is an incomplete venturi flume structure. The structure was built to divert a proportion of flow westwards across the Imera plain during large flood events in order to reduce flooding in Licata. However the channel was not finished, leaving the structure and a small area of floodplain storage. The river width in this part of the basin varies between approximately 20 and 150 m, with a mean width of approximately 90 m. Topographic data for the basin has been collected at a resolution of 2 m by airborne LiDAR flying at an altitude of 2800 m. The surface elevation measurements are subject to a vertical accuracy of ± 0.3 m with a maximum vertical error of ± 0.75 m and a horizontal accuracy of ± 0.3 m. Although there are urban regions along the coastal locations, the basin is predominantly rural.

The flood event that will be used as a case study occurred on 12 October 1991, where 229 mm of rain fell during 21 hours with a maximum intensity of 56 mm h^{-1} (Aronica *et al.*, 1998), producing widespread flooding and damage within the Imera basin. Data collected for this large flood were described by Aronica *et al.* (1998).

Unfortunately the gauge to the north of the venturi flume was destroyed during the flood; however, a previous study of this flood event used a hydrograph that was reproduced from a rainfall runoff model calibrated on previous rainfall events (Aronica *et al.*, 2002). This hydrograph, shown in Figure 2, is also utilized for this study. The validation data available for evaluating model performance include a map outlining the extent of the flood and 25 measurements of water depth taken throughout the basin. Although there was no remotely sensed data available to determine the full extent of the flood, field surveys and damage reimbursement data were used to delineate the flood boundary (Aronica *et al.*, 1998).

MODEL DESCRIPTION—LISFLOOD-FP

Models of reduced complexity were initially developed as simpler and more computationally efficient alternatives to solving the full 2D shallow wave equations. However, improvements in the computational efficiency of these models, for example by Bates *et al.* (2010), have increased

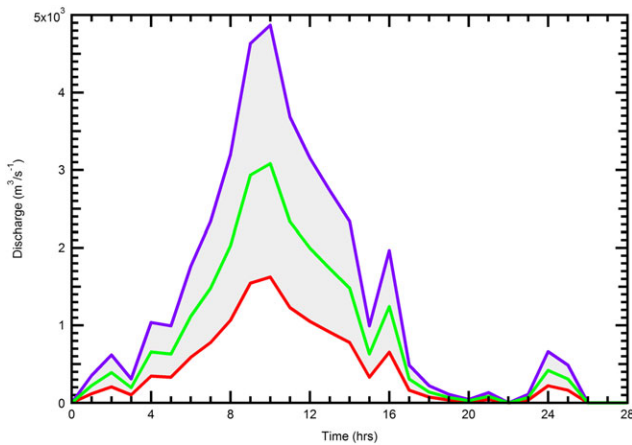


Figure 2. Hydrograph reconstructed for the 1991 flood event using rainfall runoff modelling (Aronica *et al.*, 1998). The green line in the middle of the graph is the base hydrograph; the purple and red lines are the maximum and minimum perturbations respectively. The remaining perturbed hydrographs all fall within the shaded region

the feasibility of performing probabilistic flood risk analysis at fine scale resolutions.

The hydraulic model that will be utilized here is the 2D version of LISFLOOD-FP that solves an inertial formulation of the shallow water equations (Bates *et al.*, 2010). This model is a finite difference model that is explicit in time and first order in space. LISFLOOD-FP was designed to be able to use high-resolution DEMs and so initial developments attempted to simulate dynamic flood events using a simple physical representation of the channel and floodplain processes (Bates and De Roo, 2000).

There have been multiple improvements to LISFLOOD-FP since its inception (Bates and De Roo, 2000). For example, Hunter *et al.* (2005) introduced an adaptive time step version to prevent the model becoming unstable; however, this resulted in a significant additional computational cost (Hunter *et al.*, 2006). This led to the most recent major development of LISFLOOD-FP where the diffusion wave approximation was replaced by a simplified version of the shallow water equations incorporating inertial terms (Bates *et al.*, 2010), a development that substantially improved the computational efficiency of running hydraulic simulations. The model decouples x and y calculations to allow the 1D shallow water equations that incorporate inertia to be solved at the boundary between each cell on a raster grid, allowing a 2D solution to be produced. The equation that Bates *et al.* (2010) derived to calculate the amount of water that flows between cells is:

$$Q^{t+\Delta t} = \frac{q^t - g h_{flow}^t \Delta t \frac{\Delta(h^t+z)}{\Delta x}}{\left(1 + g \Delta t n^2 |q^t| / \left(h_{flow}^t\right)^{7/3}\right)} \Delta x \quad (1)$$

where Q is flow (m^3s^{-1}), g is acceleration because of gravity (ms^{-2}), h is depth (m), n is the Manning's coefficient of roughness ($\text{sm}^{1/3}$), q is water flux (m^2s^{-1}), t is time, Δx is cell resolution (m), z is cell elevation (m) and h_{flow}^t is the depth that water can flow through between cells (m). h_{flow}^t is calculated as the difference between the highest bed elevation and the highest water surface elevation (WSE) between two cells.

Incorporating inertial terms without a diffusive term can cause instabilities for simulations with low Manning's n roughness coefficients (Bates *et al.*, 2010). To stabilize the solution, diffusion can now be introduced in a small and controlled way by using the diffusive term (θ) described by de Almeida *et al.* (2012). This diffusive term (θ) will be applied to models run with small friction coefficients that become unstable in order to stabilize their solutions.

MODELLING STRATEGY

The 2 m LiDAR data was resampled using the nearest neighbour method to create 12 DEMs with 10, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450 and 500 m resolutions (Figure 3). This method, which has also been applied previously (Fewtrell *et al.*, 2008; Neal *et al.*, 2011), allows original elevation values to be retained at nodal points and avoids smoothing effects introduced when resampling the DEM using other methods e.g. taking the mean of the elevation points. As shown in Table I, the time taken for a model simulation to run for the Imera basin at the spatial resolutions used in this study varies from 10's of minutes to seconds. This demonstrates how running models at coarse resolutions can be significantly less computationally expensive than running them at fine resolutions.

At resolutions that are coarser than the channel, the representation of the channel within the DEM will also be degraded. This means that any observed changes in model performance would be a result of the poor representation of both the floodplain and channel at coarser resolutions. The venturi flume is represented at models coarser than 20 m (the approximate width of the flume) by introducing the following weir equation which allows flow to be restricted through the cell containing the structure in a consistent and grid scale independent manner:

$$Q = CL(2gH)^{1.5} \quad (2)$$

where C is coefficient of the weir flow (defined as 1.4), g is acceleration because of gravity (m^2s^{-2}), H is water depth (m) at the upstream of the weir (representing energy head), L is width (m) of the weir (set to 20 m) and Q is flow (m^3s^{-1}). For the 10 and 20 m models, the venturi-flume width is characterized within the DEM.

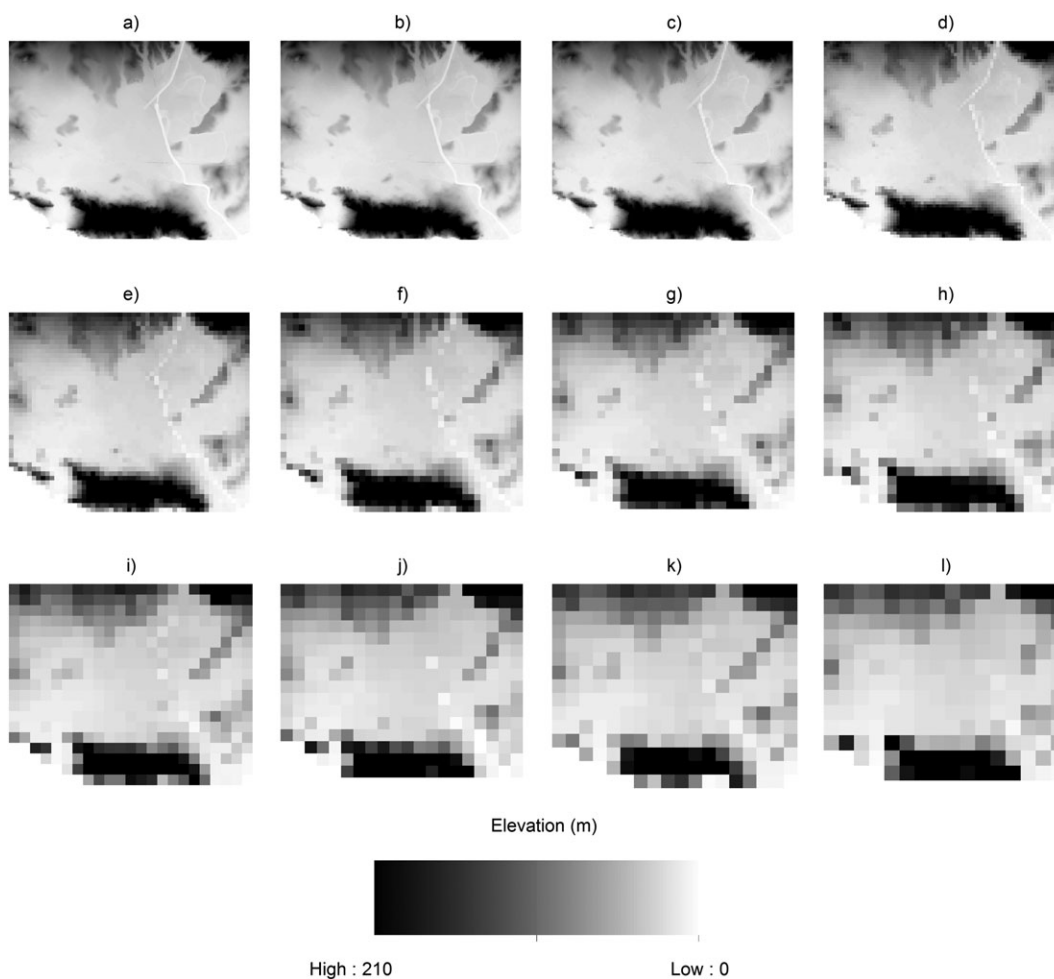


Figure 3. Digital Elevation Models (DEMs) for the a) 10, b) 20, c) 50, d) 100, e) 150, f) 200, g) 250, h) 300, i) 350, j) 400, k) 450 and l) 500 m model resolutions

Table I. Computation time for LISFLOOD-FP simulations at different spatial resolutions

| Model resolution | 10 | 20 | 50 | 100 | 150 | 200 | 250 | 300 | 350 | 400 | 450 | 500 |
|---------------------------|---------|---------|--------|------|------|-------|-------|-------|-------|-------|-------|-------|
| Simulation run time (min) | 50.98 | 6.00 | 0.52 | 0.13 | 0.12 | <0.10 | <0.10 | <0.10 | <0.10 | <0.10 | <0.10 | <0.10 |
| Number of model cells | 523 110 | 130 613 | 20 988 | 5214 | 2332 | 1320 | 832 | 572 | 437 | 320 | 270 | 208 |

Simulations computed on an Intel 3.1 GHz quad-core i7-3770S CPU.

Aricò *et al.* (2011) discuss how hydraulic models without inertia have a low sensitivity to localized narrowing of flow. However a previous study by Luke *et al.* (2015) who use the inertial formulation of LISFLOOD-FP found that although there are some local scale inaccuracies associated with flow through a narrowing location (in their case a defence breach), this had a minimal impact on the flood dynamics as a whole. Therefore although we do not model the venturi flume explicitly, representing the narrowing of flow consistently at each resolution should be adequate for the aims of this study.

Coarsening a DEM causes small scale topographic features to be smoothed over and this loss of detail could represent a key control on model behaviour at coarse scale grid resolutions. In this study, a 2D model is used with the channel and topography represented by the DEM. This means that as the DEM resolution becomes coarser than the channel, the channel become less well defined which could alter the flow dynamics of the model.

To check whether changes in model outputs at different resolutions are a result of a loss of topographic detail as opposed to inherent differences in the numerical

solution of the model at different grid resolutions, models run using coarse resolution DEMs were compared to identical model simulations with the same underlying topography but with calculations performed on a 10 m grid. It was found that the numerical differences introduced by performing calculations on coarser grids have a negligible impact on the maximum water depths predicted. Importantly, this means that any differences observed between coarse and fine models will predominantly be a result of the changing topography rather than numerical differences introduced by performing calculations with a different Δx term in Equation (1).

Discharge measurements have been shown to be uncertain by at least 40% because of both measurement uncertainties and rating curve configuration (Di Baldassarre and Montanari, 2009; McMillan *et al.*, 2012). However, the lack of gauged data available for this flood event means there is limited evidence with which to be objective about the true uncertainties. Therefore, as the flood was extremely large, the uncertainty of the hydrograph was represented by making perturbations at increments of 5% between 50% and 150% to create 21 possible inflow hydrographs that are given equal weighting. The shape of the hydrographs remains the same for each realization. This could be treated as a possible conservative scenario of the likely errors given evidence constraints and the magnitude of the event.

To represent the uncertainty introduced through the friction parameters required by the model, simulations are run with varying friction parameters. In this study the friction parameter is spatially lumped with separate values for the channel and floodplain. These friction values are sampled systematically with an interval of 0.01, with the channel friction coefficients varied between 0.02 and 0.1 and the floodplain friction coefficients varied between 0.02 and 0.2. This results in 171 possible combinations of friction parameters. These parameter ranges were used for a previous study on this catchment (Aronica *et al.*, 2002) and are wide enough to explore whether the parameter values that give optimum model performance are non-stationary with respect to spatial resolution changes (Fewtrell *et al.*, 2008) and to different magnitude flows (Cloke *et al.*, 2013).

Models are simulated for each possible parameter combination and each inflow perturbation. For each spatial resolution there are therefore a total of 3591 simulations. Initially all models were simulated without the introduction of the diffusive term (θ); however, the term was introduced with a value of 0.9 for the 10 and 20 m models that had a channel friction coefficient of 0.02 as these simulations produced large mass errors as a result of instabilities. Subsequently, the mass error for all simulations is negligible.

Quantifying validation data uncertainty

Like all hydrological data (McMillan *et al.*, 2012), uncertainties contained within the observational data used to constrain model performance in this study are subject to error. However, by combining newer data such as high-resolution LiDAR with older observations it is possible to quantify likely error bounds of these data, allowing a fair assessment of model performance to be made. Intersecting the 2 m Lidar data with the inundation shoreline that was delineated post event (Aronica *et al.*, 2002) revealed that the gradient of WSE along some sections of the shoreline was implausibly steep. Given that the use of WSEs to assess model performance has been shown to allow improved diagnostic analysis (Mason *et al.*, 2009), it was decided to use the water depth observations to rank competing models.

It is important that the uncertainty of the observational data is accounted for in any analysis. For observations of water depth, uncertainty consists of both horizontal and vertical components. The horizontal uncertainty is the uncertainty of the measurement location, whilst the vertical uncertainty is the uncertainty of the measurement itself and of the mark representing a maximum water depth. For example, capillary action or ponding of flood waters could lead to water marks that are misrepresentative of the actual maximum water depth at a location.

The location of the observational water depth data for the Imera flood event was determined using a map rather than by GPS meaning that the exact location of the measurement is uncertain. To assess this uncertainty, the data were intersected with the 2 m LiDAR data and a structure overlay to determine possible locations where the water depth marks were measured. Maximum and minimum horizontal uncertainty bounds of the measured WSE were then calculated by adding possible surface elevations to water depth measurements.

The vertical uncertainty of water depth measurements has previously been estimated to be in the region of ± 0.5 m (Fewtrell *et al.*, 2011b). This was for an urban flood event in Carlisle where wrack and water marks were measured using differential Global Positioning Systems (GPS). The small sample size of observational data for the Imera flood event makes it difficult to quantify the vertical uncertainty of the observations; therefore, this study will use an approximate estimation of the vertical uncertainty to be ± 0.5 m. The horizontal, vertical and LiDAR uncertainty (± 0.30 m) were combined to construct the upper and lower uncertainty ranges of the WSE at each of the locations where observational water depth data exists.

Performance metric

Epistemic errors are rarely unstructured errors of normal distribution with a zero mean (Beven *et al.*, 2011b),

and given the uncertainties discussed previously it is very unlikely that the WSE errors described in this study would meet the underlying assumptions required by commonly used metrics, for example the root mean squared error (RMSE) metric. In hydrology there are a number of approaches that have used fuzzy metrics that allow the inclusion of observational uncertainties into the assessment of model performance (Siebert and Mcdonnell, 2002; Freer *et al.*, 2004; Beven, 2006). These methods have advantages for defining whether a model simulation is behavioural or not as model predictions that do not fit within defined limits of acceptability are classified as non-behavioural. However although, we have used the best data available to encapsulate the observational uncertainties, we cannot be 100% confident that the bounds represent the absolute limits of uncertainty. This means that it would be unfair to require a model to fit all of the observational points within the ranges described above (i.e. to employ strict limits within our acceptability framework).

The uncertainties used to construct the WSE uncertainty bounds are independent of one another; hence, treating these uncertainties as additive errors would represent a worst case scenario. It is unlikely that the true error of a WSE observation will be constructed at the limits of each of these uncertainties at the same time. We have therefore used a performance metric based on a monotonically increasing triangular function whereby a model is given a score for each of the observed WSE observations such that the model scores 0 if the predicted WSE is at the mid-point of the observational uncertainty bounds, +1 if the predicted WSE is equal to the upper bound and -1 if the predicted WSE is equal to the lower bound:

$$Point\ score = \frac{Modelled\ WSE - Observed\ WSE\ mid-point}{0.5(Observed\ WSE\ upper\ bound - Observed\ WSE\ lower\ bound)} \tag{3}$$

This function monotonically increases linearly so that if a modelled WSE is equal to two times the difference between the uncertainty bound and mid-point it will score ± 2 . The mean of the absolute scores for each observation

is then taken as the overall performance score for a model simulation:

$$Model\ simulation\ score = \frac{\sum_{i=1:n} \sqrt{Point\ score_i^2}}{n} \tag{4}$$

where n is the number of observations. If a model predicts that there is no water in a location where there was an observation, the WSE was taken from the nearest cell where water was present, using the method described by Neal *et al.* (2009). Using this approach avoids the model being given a score limited by the difference between the lower bound and the surface elevation at an observation location if no water is predicted at that location, which would make the metric insensitive when the model significantly underestimates flood extent. At coarser resolutions where two or more observations occasionally fall into the same grid cell, the range for the horizontal uncertainty is obtained by taking the maximum and minimum WSEs from these observations and combining

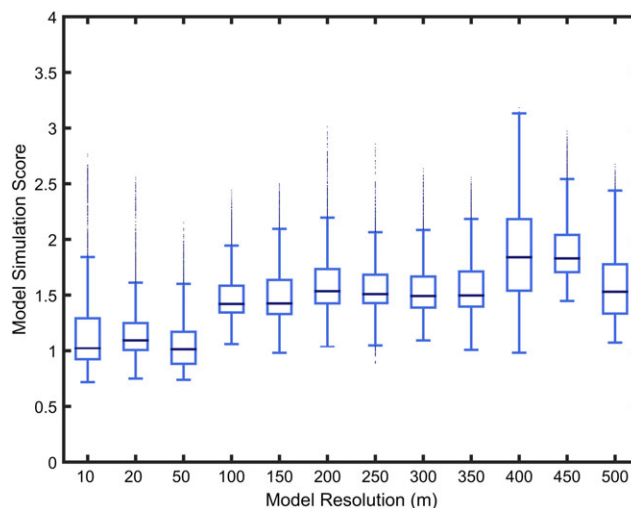


Figure 4. Boxplots of model performance grouped by model resolution. The centre of the boxes represents the 50th percentile of the range of simulation scores for each model resolution. The top and bottom of the boxes represent the 75th and 25th percentiles so that the scores in the box represent the inter quartile range (IQR). The whiskers of the boxes contain any other model simulation scores that are greater or less than those contained within the boxes up to a maximum of distance of 1.5 times the IQR away from the nearest quartile. Any scores exceeding this distance away from the box are classified as an outlier and are represented as a dot on the graph

Table II. Model Performance for each resolution using uncertain boundary conditions and model parameters

| Resolution (m) | 10 | 20 | 50 | 100 | 150 | 200 | 250 | 300 | 350 | 400 | 450 | 500 |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Minimum fuzzy score | 0.72 | 0.75 | 0.74 | 1.06 | 0.98 | 1.03 | 0.89 | 1.09 | 1.01 | 0.98 | 1.45 | 1.07 |
| Q5 fuzzy score | 0.82 | 0.87 | 0.79 | 1.26 | 1.23 | 1.28 | 1.31 | 1.25 | 1.33 | 1.15 | 1.58 | 1.15 |
| Q25 fuzzy score | 0.92 | 1.01 | 0.88 | 1.34 | 1.33 | 1.43 | 1.43 | 1.39 | 1.40 | 1.54 | 1.71 | 1.33 |
| Q50 fuzzy score | 1.03 | 1.09 | 1.01 | 1.42 | 1.42 | 1.54 | 1.51 | 1.49 | 1.50 | 1.84 | 1.83 | 1.53 |
| Q75 fuzzy score | 1.29 | 1.25 | 1.17 | 1.58 | 1.64 | 1.73 | 1.68 | 1.67 | 1.71 | 2.18 | 2.04 | 1.78 |
| Q95 fuzzy score | 2.13 | 1.78 | 1.57 | 2.01 | 2.09 | 2.24 | 2.10 | 2.14 | 2.16 | 2.72 | 2.52 | 2.21 |
| Maximum fuzzy score | 2.76 | 2.56 | 2.15 | 2.44 | 2.50 | 3.01 | 2.86 | 2.64 | 2.56 | 3.18 | 2.97 | 2.67 |

them to give a wider range from which the mid-point for this set of observations is taken. This means that coarse models are not unfairly penalized by having to fit two separate observations from one model grid cell prediction.

Removing disinformative data

Even after the error characterization procedures identified above, it can still be the case that patterns in the observational data are not what would be expected. For example, erroneously large gradients between maximum WSEs could indicate that some of the observational data may be disinformative. Such behaviour is physically unrealistic over such small distances and consequently should be rejected. To identify an observation as erroneous the difference between the WSE ranges at this location need to be significantly larger or smaller than nearby observations, such that the water surface slope between two observations is implausible. The minimum difference in WSE was calculated for each possible pair of observations, given their associated uncertainty bounds. These differences were then split into bins based on the distances between the two observations and the

mean taken for each 200 m distance and plotted. The gradient through this line represents an estimate of what the mean observed water surface slope was for this flood event. If the water surface slope between two observations is greater than two times the gradient then the observations were highlighted as potentially disinformative. Consequently, observations 1, 2, 8, 9, 10, 12 and 14 were classified as erroneous and removed from the analysis, leaving 18 observations from which model performance was evaluated against.

Reliability of probabilistic inundation maps

The performance metric described above will allow the assessment of how the performance of individual simulations changes with spatial resolution. However this alone does not allow the assessment of the subsequent impact on the skill of the conditional probabilistic inundation maps produced. One method utilized by Horritt (2006) and Stephens and Bates (2015) that allows a comparison between probabilistic inundation maps is to calculate the reliability of these maps. Horritt (2006) does this by aggregating the uncertainty in the inundation maps

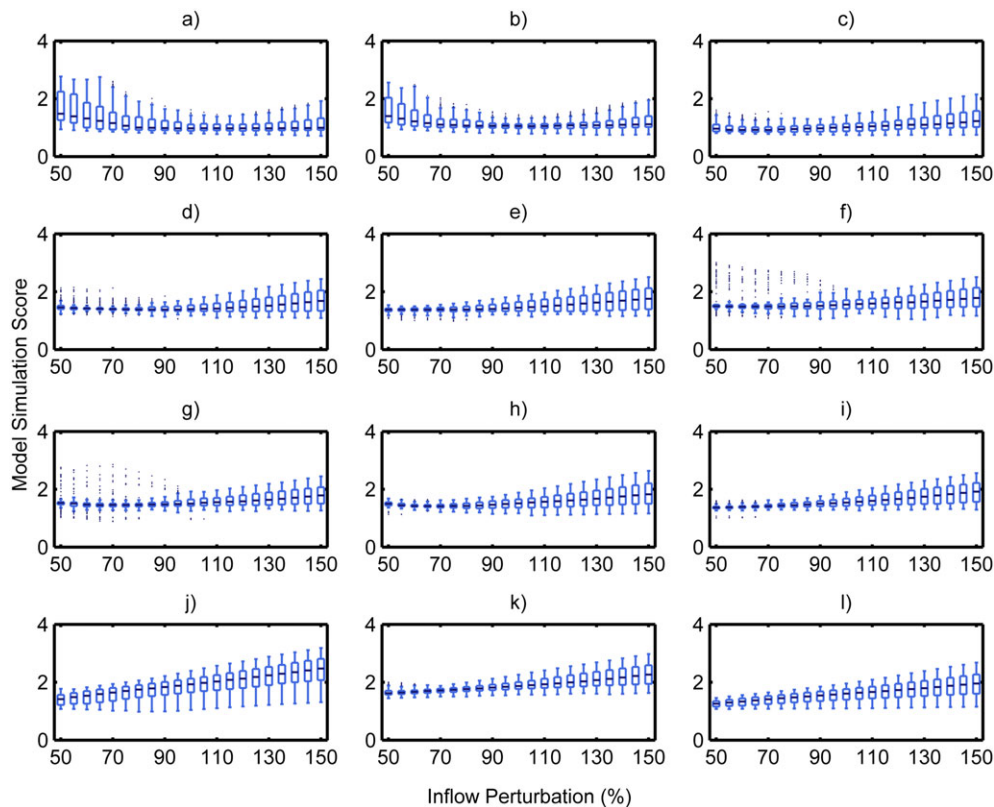


Figure 5. Boxplots of model performance grouped by inflow boundary condition perturbation for the a) 10, b) 20, c) 50, d) 100, e) 150, f) 200, g) 250, h) 300, i) 350, j) 400, k) 450 and l) 500 m model resolutions. The centre of the boxes represents the 50th percentile of the range of simulation scores for each model resolution. The top and bottom of the boxes represent the 75th and 25th percentiles so that the scores in the box represent the inter quartile range (IQR). The whiskers of the boxes contain any other model simulation scores that are greater or less than those contained within the boxes up to a maximum of distance of 1.5 times the IQR away from the nearest quartile. Any scores exceeding this distance away from the box are classified as an outlier and are represented as a dot on the graph

and the observed flood extent to assess whether the uncertainty predicted reflects the proportion of cells inundated in an area. For example if the probabilistic inundation map predicts that there is an 80% chance of flooding, then to be reliable 80% of the cells should be observed as flooded. Stephens and Bates (2015) use a similar approach but using predictions of WSEs. Here they take the Probability Distribution Function (PDF) of the predicted WSEs at each observation location and assess where in the PDF the observation lies. They split the PDFs into bins every 10% and, for a reliable probabilistic prediction, would expect 10% of the observations to fall within the 0–10% bin, 10% of the observations to fall in the 10–20% bin and so forth. In this study we assess model performance on WSEs rather than flood extent and have therefore applied an adapted version of the methodology presented by Stephens and Bates (2015) to assess the reliability of our inundation maps. Our approach differs to Stephens and Bates (2015) in that we account for the uncertainty in our observations in our cumulative frequency plots and use larger bin sizes.

First the distribution of predicted WSEs is taken for each observation at each resolution. Then the location of the upper and lower bounds and the mid-point of the observed WSE in the probability distribution is determined at each location. The probability distribution is

split into five bins: 0–20%, 20–40%, 40–60%, 60–80% and 80–100%. For each of the upper bound, lower bound and mid-point, the number of observations that fall into each bin is determined. If the predictions of WSE are perfectly reliable we would expect 20% of observations to fall into each bin and would expect the uncertainty plume to fully encapsulate the 1:1 line of the cumulative frequency plot produced from this analysis. The location of the uncertainty bounds on the y axis at $x=0$ indicates the percentage of observations that fall outside the lower bound of the WSE probability distribution function (Stephens and Bates, 2015). Subtracting the y axis value of the uncertainty bound from 100 at $x=100$ indicates the percentage of observations that fall outside the upper bound of the WSE probability distribution function (Stephens and Bates, 2015). We will assess the reliability of the conditional probability of inundation maps produced for each spatial resolution.

RESULTS AND DISCUSSION

Model performance and sensitivities

The best performing simulations across all 3591 simulations, encapsulating uncertainty in the boundary conditions and parameters, occur for the 10, 20 and 50 m

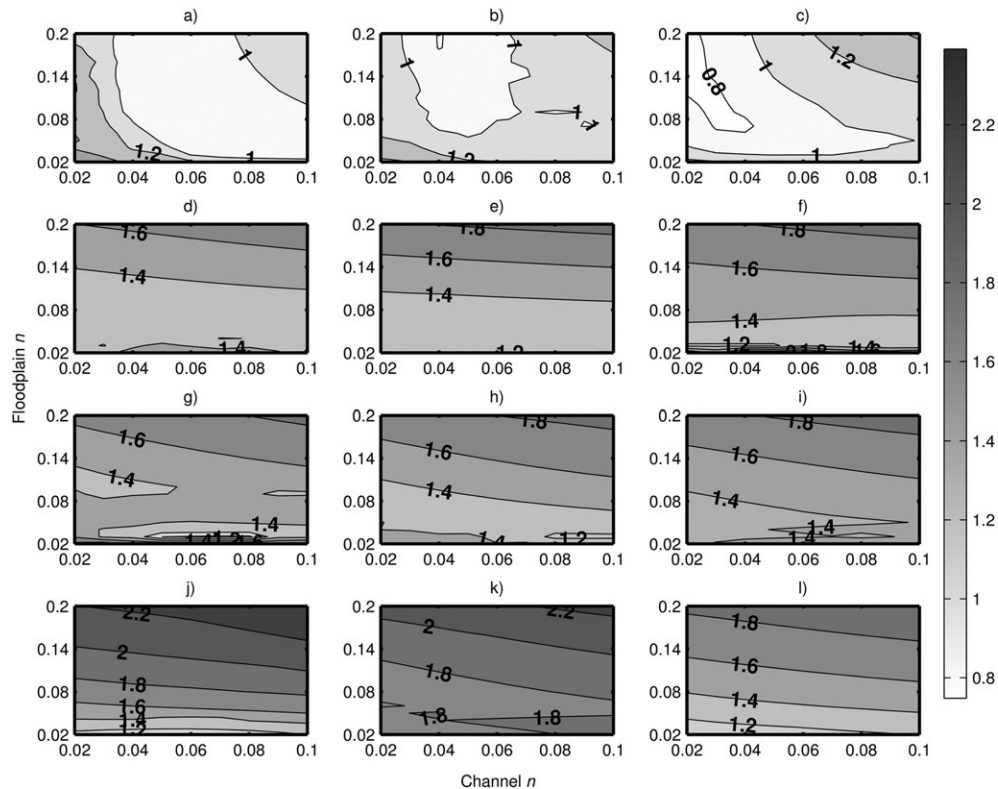


Figure 6. Model simulation score for unperturbed inflows at the a) 10, b) 20, c) 50, d) 100, e) 150, f) 200, g) 250, h) 300, i) 350, j) 400, k) 450 and l) 500 m model resolutions. A lower score and lighter colour indicate better performance, whilst a higher score and darker colour indicate worse performance

resolution models (Table II). Model performance decreases and the inter quartile range (IQR) shifts upwards as resolution coarsens beyond 50 m (Figure 4). Despite an increase in computational cost by over an order of magnitude (Table I), there is very little performance gain when running models with resolutions finer than 50 m. When breaking down model performance for each resolution by boundary condition perturbation (Figure 5) it can be seen that for the finer resolutions, the greatest variance of model performance occurs for the extreme perturbations. However, as the resolution coarsens the region of least variation, demonstrated by the small IQR, shifts towards the smallest inflows (i.e. 50% perturbation). Furthermore, model performance for our best estimate of discharge is greatest at the finest resolution and as the model resolution degrades, the models produce a poorer set of simulations. This demonstrates that at coarser resolutions, the loss of topographic detail influences the model's sensitivity to boundary condition which is non-stationary across different model resolutions.

As shown in Figure 6 for unperturbed flows, the model sensitivity to friction parameters is also non-stationary

with respect to model resolution. At fine resolutions the model is predominantly sensitive to channel friction but becomes sensitive to floodplain friction at coarser resolutions. It is also noticeable that only the 10 to 50 m resolution models have a relatively L-shaped sensitivity response pattern that is typical of two parameter hydraulic models (Horritt and Bates, 2001b; Aronica *et al.*, 2002; Werner *et al.*, 2005; Jung *et al.*, 2012) and caused by trade-offs between the channel and floodplain friction coefficients.

The reasons behind the marked shift in model performance and changes in parameter sensitivity for models coarser than 50 m can be explored further by comparing the predicted WSEs at the observation locations for the different spatial resolutions. For the majority of observations, the finest resolution models tend to predict lower WSEs and the coarsest models tend to predict the highest (Figure 7). The fact that these changes in performance and sensitivity occur after the spatial resolution is degraded to resolutions coarser than 50 m corresponds with the resolution beyond which the channel can no longer be adequately represented within the DEM.

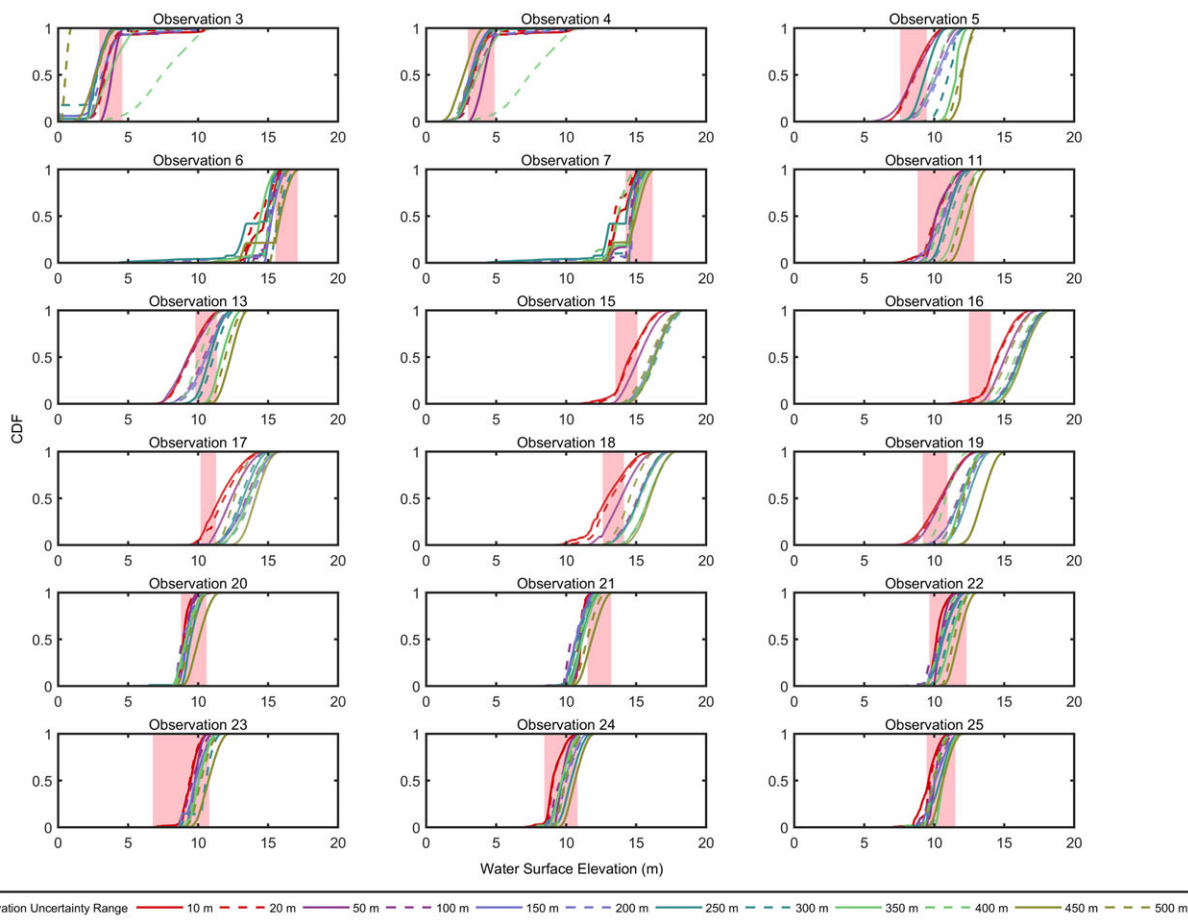


Figure 7. Cumulative distribution function (CDF) plots showing predicted WSEs for each model resolution. The shaded area in each graph is the uncertainty associated with that WSE observation

This signifies that the degradation of the channel at coarser resolutions causes a greater proportion of flow to spill onto the floodplain than in the finer resolution models, increasing the modelled floodplain water depths and velocities. The increase in water depths leads to poorer model performance, whilst the increase in floodplain flow velocity increases the sensitivity to the floodplain friction coefficient. This is because frictional force is proportional to the square of velocity multiplied by the Manning's friction coefficient.

The increase in floodplain water depths at coarser resolutions caused by the poorer channel representation also explains why the coarser models tend to perform better for lower flows (Figure 5). At the unperturbed or larger flows there is too much water spilling onto the floodplain which leads to poor performance metric scores

as the model is over predicting the WSE at observation locations. However, as this is not a problem for the finer resolution models where the channel is adequately represented, the model performance at fine resolutions is better for our best estimate of discharge. Degrading the spatial resolution of topography can therefore have a major control on the behaviour and performance of hydraulic models particularly when the channel is defined within the DEM.

If we fix the spatial resolution at the finest resolution (10 m), we find that model parameters are also sensitive to the boundary condition perturbation (Figure 8), with the best performance for lower flows requiring higher channel friction coefficients and the best performance for higher flows requiring lower floodplain or channel friction coefficients. This is because at low flows, the model

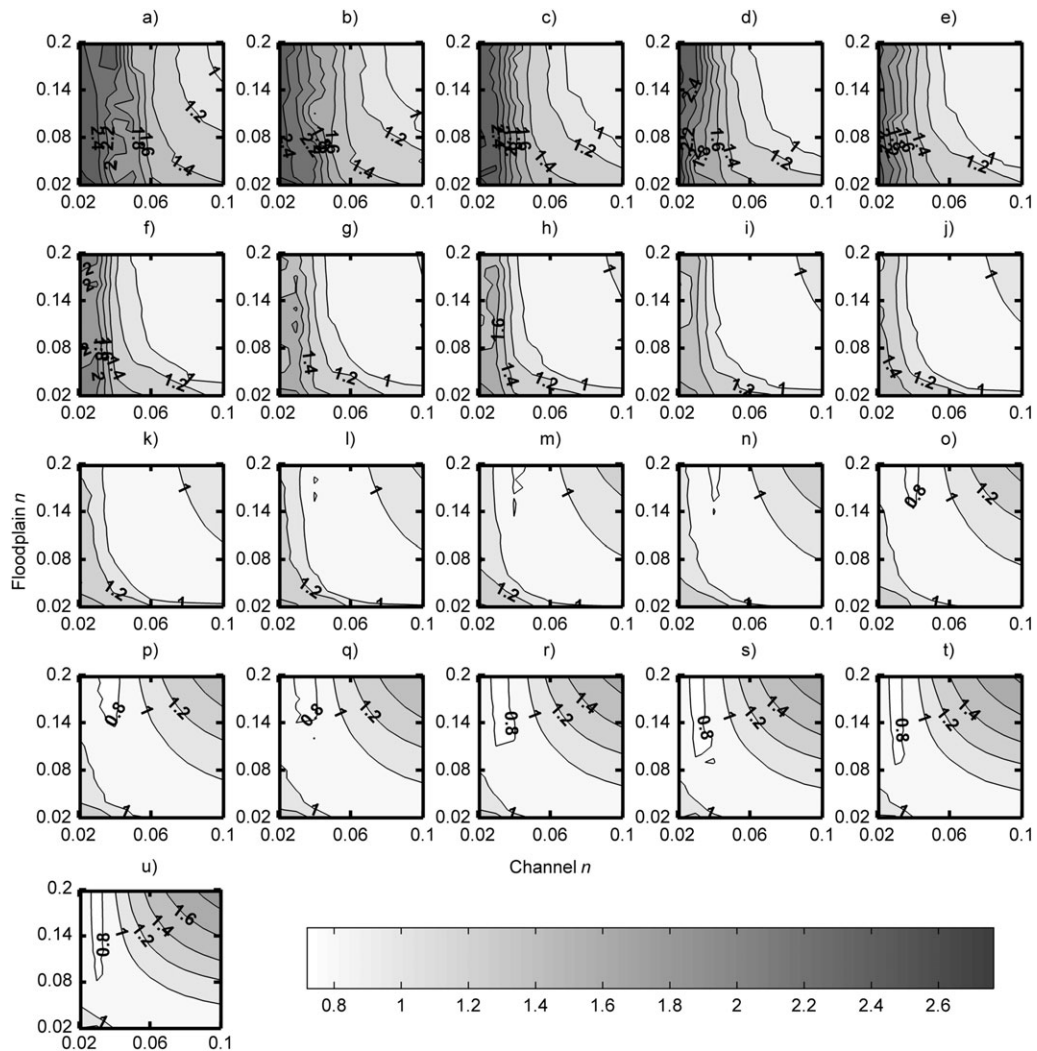


Figure 8. Model simulation score for the 10 m models with boundary condition perturbations of the inflow hydrographs ranging from a) 50% to u) 150% of the base hydrograph. A lower score and lighter colour indicate better performance, whilst a higher score and darker colour indicate worse performance. As the inflow perturbation increases through a–u the area of optimal model performance shifts from the top right of the plot to the left. The model remains broadly the same sensitivity to the friction parameters for most perturbations except for the extreme small perturbations where the model becomes more sensitive at lower friction coefficients

needs to slow the channel flow with high friction coefficients to force more of the water onto the floodplain, whilst at high flows, the friction coefficient needs to be lower for both the channel and floodplain to increase the propagation speed of the flood wave to reduce the peak water depths. Despite this compensating behaviour, the shape of the sensitivity plots remains broadly similar suggesting that the model's sensitivity to friction parameters remains constant for a given model resolution despite different magnitude events. This also demonstrates that the parameter sensitivities vary in a consistent manner that reflects their physical basis.

However, the fact that the model can compensate for larger and smaller inflow perturbations and still produce similar levels of performance clearly demonstrates that calibrated friction parameters are completely effective and can subsume errors in the boundary conditions. Therefore it is important to collect improved validation data to help to constrain the number models classified as behavioural and therefore allow the uncertainty that propagates into probabilistic flood inundation maps to be reduced. In particular, time stamped water depth observations would be extremely beneficial to condition model performance

and reduce the model's ability to compensate for large and small flows in this way.

Deterministic versus probabilistic flood maps

Although the changing skill of a model can be assessed by analysing performance metrics, maps of predicted spatial inundation are useful tools to understand how model behaviour changes at different resolutions. To produce conditional probability of inundation maps, we use a method utilized in previous studies (for example Aronica *et al.*, 2002) where model predictions of maximum water depth are converted into a binary index. A cell is classified as 1 if the cell is wet (defined as a maximum water depth equal to or greater than 0.10 m) and 0 if the cell is dry (defined as a maximum water depth of less than 0.10 m). Each model is given a weighting, defined as the inverse of the performance metric, meaning that the best performing models are given a greater weighting. The weighted model outputs are then combined to produce conditional probability of inundation maps where a cell with a score of 1 indicates a 100% likelihood of flooding, 0 indicates 0% likelihood of

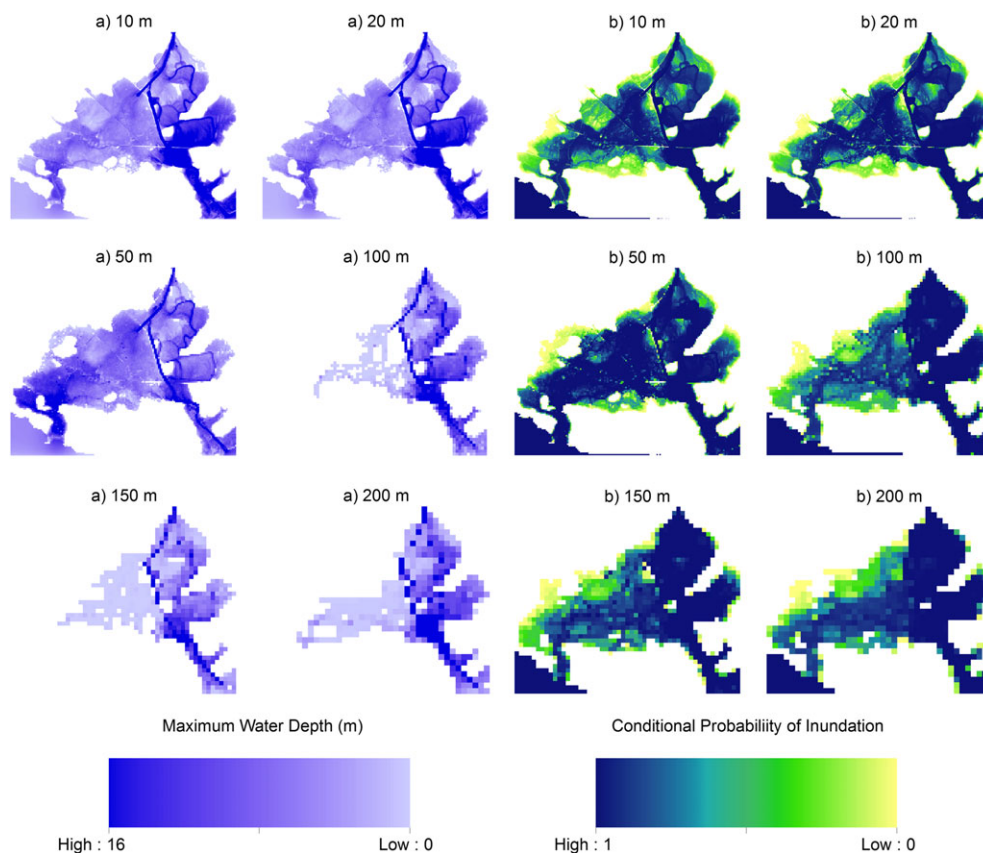


Figure 9. a) Deterministic maximum water depth and b) conditional probability of inundation maps for the 10, 20, 50, 100, 150 and 200 m model resolutions. A conditional probability of 0 represents a 0% chance of inundation in a cell, whilst a conditional probability of 1 represents a 100% chance of inundation in a cell. A conditional probability of 50% indicates the greatest uncertainty

flooding and 0.5 (or 50% probability) represents the maximum uncertainty.

The conditional probability of inundation maps are presented alongside deterministic inundation maps (Figures 9 and 10), defined as the best performing simulation for each model resolution. From these maps it can be seen that for each model resolution there are areas within the flood extent where flooding is uncertain. Furthermore, there are areas where the 10 m deterministic model has precisely identified regions that will not flood but where the probabilistic models have identified a possibility of flooding. This shows that using a deterministic flood maps to inform decisions would be misleading and not representative of the overall uncertainties that are inherent in making these inundation predictions. At fine resolutions the added precision of the deterministic flood maps could lead to increased confidence in their detail and subsequently lead to spurious decision making.

The deterministic and probabilistic flood extent boundaries are broadly similar for the finer resolution models; however, it is clear from Figures 9 and 10 that the

deterministic maps appear to break after 100 m where for some of the coarser models the deterministic solutions fail to reproduce the flooding to the west of the domain. Despite this the probabilistic maps for even the coarsest resolutions are still able to produce a coherent flood outline that is comparable to the finer resolution models. Some of the probabilistic flood maps at coarser resolutions do however have different certainties of inundation to the finer resolution maps, with the coarser models tending to produce a higher certainty of inundation to the East of the domain and lower to the West which is likely to be caused by the increase in floodplain flow at coarser resolutions that was discussed previously. It is also clear that resampling topography to coarse resolutions can significantly alter model behaviour, as demonstrated by the 400 m flood maps where it is shown that the flood does not reach the sea in the South West corner of the domain.

To assess whether the skill in the inundation maps changes as resolution coarsens, reliability plots were produced (Figure 11). From Figure 11 we can see that for the 10, 20, 50 and 100 m resolutions, the 1:1 gradient line

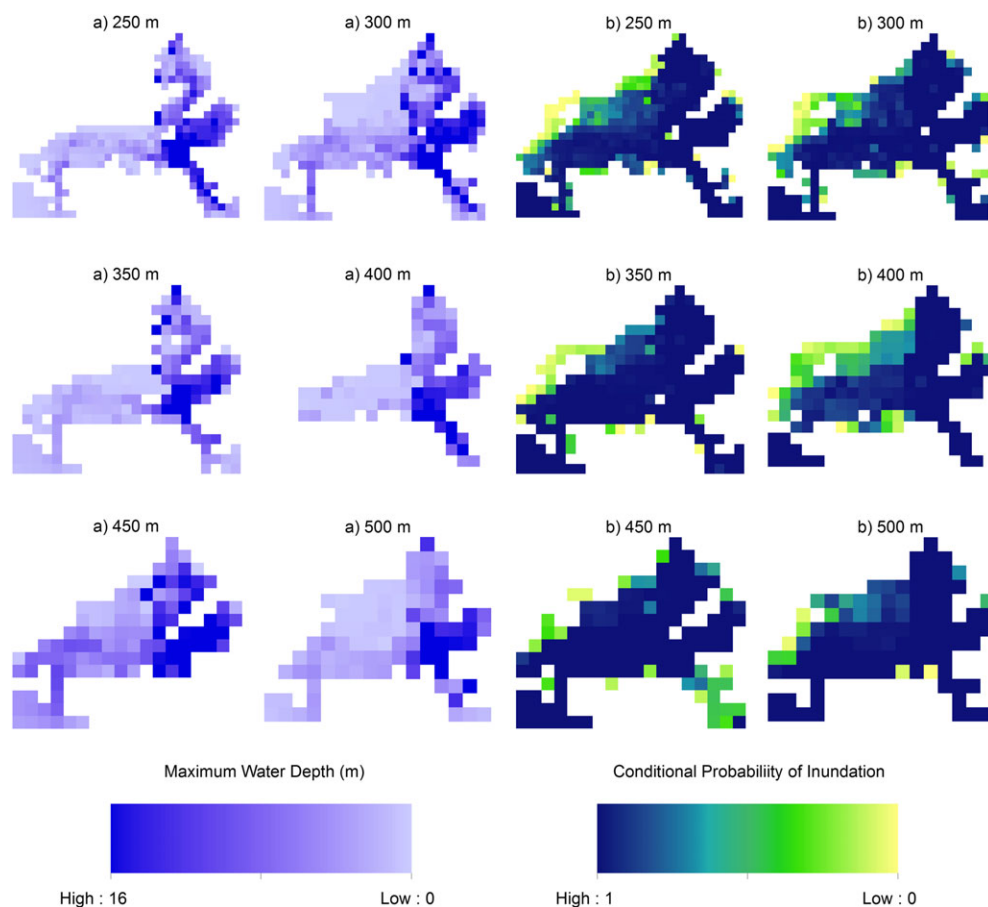


Figure 10. a) Deterministic maximum water depth and b) conditional probability of inundation maps for the 250, 300, 350, 400, 450 and 500 m model resolutions. A conditional probability of 0 represents a 0% chance of inundation in a cell, whilst a conditional probability of 1 represents a 100% chance of inundation in a cell. A conditional probability of 50% indicates the greatest uncertainty

falls within the uncertainty range of the reliability plot. This means that the observed WSEs are evenly distributed within the modelled WSE PDFs, therefore indicating the probabilistic inundation predictions at these resolutions to be reliable. However as the resolution coarsens, the reliability decreases as the 1:1 gradient line is not fully encapsulated by the uncertainty plume. Furthermore the uncertainty plume of the observed WSE percentile is less than 1 at $x=100$ for some resolutions, indicating there are observations that fall outside of the predicted PDF of the models. However this degradation is non-linear, for example the 400 m resolution model produces a more reliable probabilistic inundation map than the 350 m resolution model. This demonstrates that although the inundation extent predicted by probability of inundation maps remains broadly consistent for each spatial

resolution (Figures 9 and 10), the reliability of the uncertainties predicted within the maps are non-stationary with resolution and degrade at resolutions coarser than 100 m.

Up to this point no model simulations have been rejected as non-behavioural to allow comparisons across all spatial resolutions; however, often poorer models are rejected when employing a limits of acceptability approach within an uncertainty analysis and this can impact on subsequent predictions. Given that the uncertainty bounds of the observations are a best estimate, an appropriate threshold to accept or reject models in this study would be to classify models that have an overall simulation score 1 or below as behavioural (Figure 4) and reject models that score above 1 as non-behavioural. This means that on average the accepted models fit within the uncertainty bounds of the observed WSEs. Employing

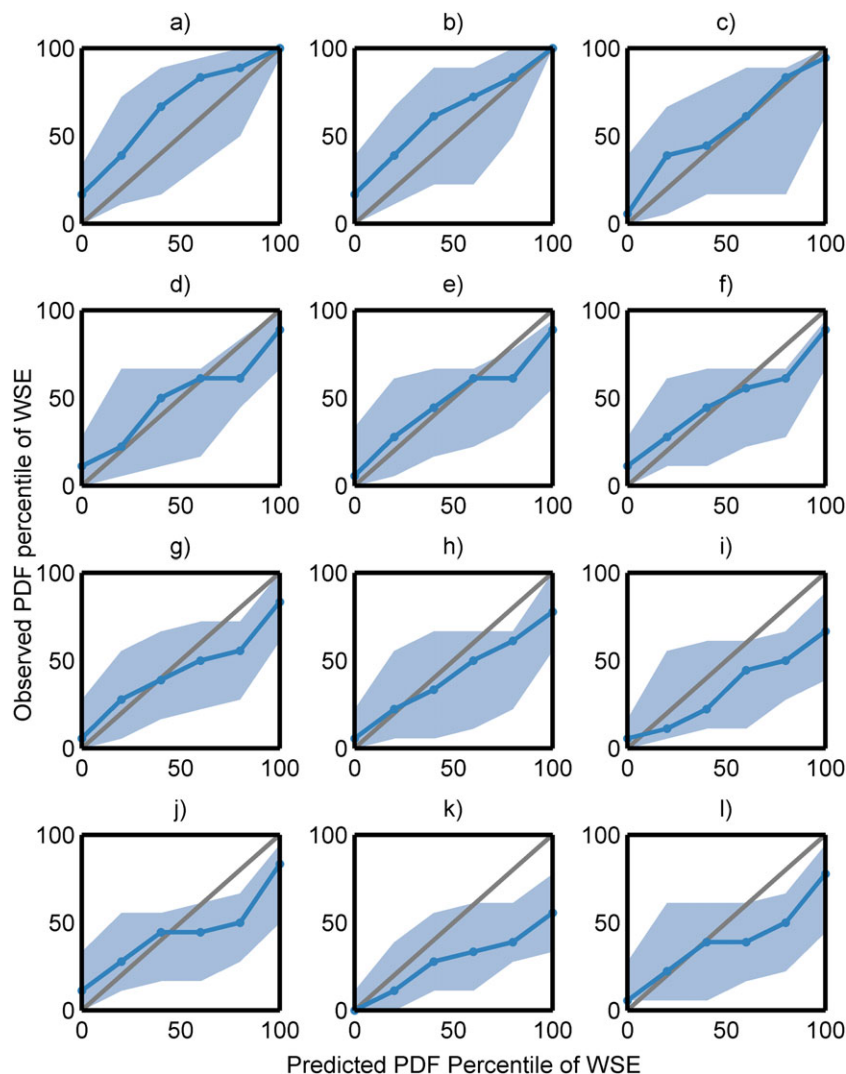


Figure 11. Reliability of probabilistic inundation maps for the a) 10, b) 20, c) 50, d) 100, e) 150, f) 200, g) 250, h) 300, i) 350, j) 400, k) 450 and l) 500 m model resolutions. The solid blue line represents the midpoint of the observed water surface elevation (WSE) uncertainty ranges, and the shaded area is calculated by taking the upper and lower uncertainty bounds of the observed WSE. For perfect reliability the 1:1 gradient shown by the solid grey line should be fully encapsulated by the shaded area

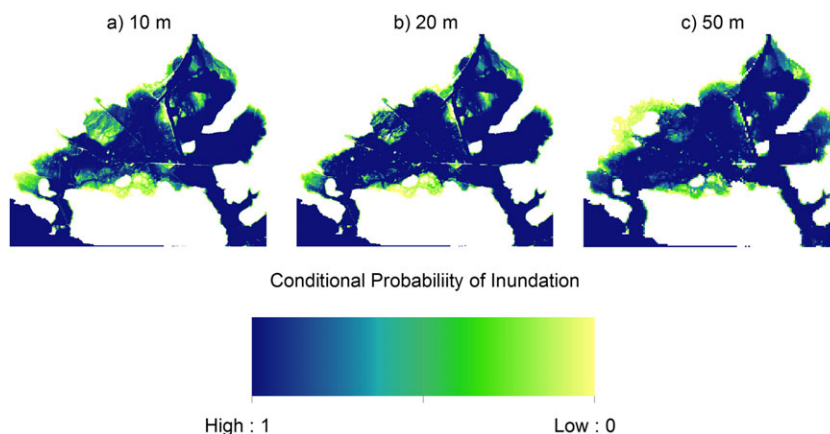


Figure 12. Conditional probability of inundation maps for the 10, 20 and 50 m model resolutions using behavioural simulations. A conditional probability of 0 represents a 0% chance of inundation in a cell, whilst a conditional probability of 1 represents a 100% chance of inundation in a cell. A conditional probability of 50% indicates the greatest uncertainty

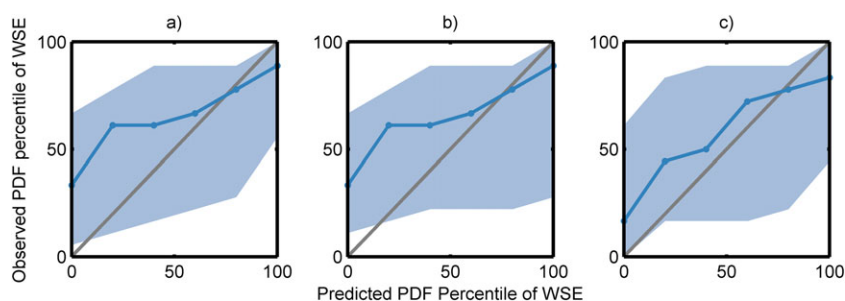


Figure 13. Reliability of the behavioural probabilistic inundation maps for the a) 10, b) 20 and c) 50 m model resolutions. The solid blue line represents the midpoint of the observed water surface elevation (WSE) uncertainty ranges, and the shaded area is calculated by taking the upper and lower uncertainty bounds of the observed WSE. For perfect reliability the 1:1 gradient shown by the solid grey line should be fully encapsulated by the shaded area

this threshold means that at resolutions coarser than 50 m, only a small amount of simulations would be classified as behavioural and at the majority of these coarser resolutions no models would be accepted. Behavioural probabilistic flood maps were therefore calculated for the 10, 20 and 50 m spatial resolutions. These maps produce very similar flood outlines to the original probabilistic maps however there are fewer regions of high uncertainty (Figure 12). This is demonstrated by an increase in the prominence of dark and light shades of a blue and a reduction of the mid-shades of blue when compared to the probabilistic flood maps where all models were accepted (Figure 9).

However, a consequence of this sample size reduction is that the probabilistic flood maps become less reliable. Figure 13 shows that for the 10 and 20 m models, there are observations that lie outside of the lower end of the predicted PDF. This shows that the subset of behavioural models at these resolutions over predict WSEs at the observation locations in comparison to when including the full range of model simulations (Figure 11). Although the 50 m reliability plot for behavioural simulations

indicates that the probabilistic map is reliable, there is a marked shift for all three resolutions where the midpoint line and uncertainty plumes of the reliability plots (Figure 13) become much shallower than when all simulations are considered (Figure 11). This indicates that fewer observations fall within the predicted PDF of WSEs, reducing the reliability of the behavioural probabilistic flood maps. Consequently, decisions made based on a probabilistic inundation map composed using a subset of possible model realizations may be mistakenly overconfident in their nature. Modellers should therefore consider carefully the process of eliminating non-behavioural model simulations.

CONCLUSIONS

In this paper we have assessed how changing the resolution of a hydraulic model with uncertain parameters and boundary conditions impacts on conditional probabilistic maps of flood inundation. We have presented a methodology that assesses the uncertainty of observational

data, removes those that are disinformative and have constructed a performance metric that accounts for these uncertainties in a consistent manner.

We have demonstrated for this study that the skill of the deterministic simulations degrades at resolutions coarser than 50 m. This is predominantly caused by the channel being poorly represented within the DEM at coarser resolutions, leading to increased floodplain water depths and flow velocities which subsequently results in poorer model performance and changes in parameter sensitivity. However, there is very little performance gained by running simulations of a large rural flood event at resolutions finer than 50 m and doing so will incur unnecessary additional computational cost.

Furthermore, it has been shown that using fine resolution models deterministically can lead to spuriously precise flood predictions that do not represent the uncertainty that is inherent in making flood inundation predictions, as demonstrated by the probabilistic flood maps. The ability to run models at coarser resolutions represents a significant computational benefit with simulation run time typically decreasing by an order of magnitude for a doubling of the resolution. This becomes important when there is limited computational resource and when multiple simulations are required for probabilistic analysis. It would therefore be more beneficial for a decision maker to spend their computational resource producing probabilistic flood maps at coarser resolutions rather than deterministic flood maps at a finer resolution. However, although the delineation of flood extent is similar for the probabilistic maps at different resolutions, the reliability of these probabilistic inundation maps deteriorates when the spatial resolution is coarsened beyond 100 m. Additionally, removing non-behavioural models reduces the uncertainty in the probabilistic flood maps; however, doing so also reduces the reliability of these maps which could lead to overconfident decision making.

Although the model's sensitivity to parameters remains broadly similar for different inflow perturbations reflecting the physical basis of the parameters, the fact that the optimum model performance in the parameter space shifts demonstrates the ability of the parameters to subsume errors in the boundary conditions. Consequently there is a need to collect high quality spatially variable and temporally defined water depth observational data that will help in validating hydraulic models and constraining model parameters and boundary conditions. Future work should explore whether similar conclusions are relevant in an urban environment, for example as Fewtrell *et al.* (2008) did for deterministic models.

What is clear from this study is that flood maps are uncertain, and presenting these as error-free and deterministic is highly misleading and may lead to either

over-confidence or poor decision making. We also need to continue developing improved methodologies that incorporate uncertain flood information into decision making and train a new generation of risk managers who are capable of using more sophisticated probabilistic risk information.

ACKNOWLEDGEMENTS

This work was supported by the Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1]. The authors would like to thank the Regione Siciliana—Assessorato del Territorio e dell'Ambiente—S.I.T.R., who provided the LiDAR data used in this study. The authors would also like to thank Elisabeth Stephens and an anonymous reviewer for their insightful comments and suggestions which have helped improve this manuscript.

REFERENCES

- de Almeida GAM, Bates P, Freer J, Souvignet M. 2012. Improving the stability of a simple formulation of the shallow water equations for 2-D flood modelling. *Water Resources Research* **48**: W05528.
- Apel H, Merz B, Thielen HH. 2009a. Influence of dike breaches on flood frequency estimation. *Computers & Geosciences* **35**: 907–923.
- Apel H, Aronica GT, Kreibich H, Thielen AH. 2009b. Flood risk analysis—how detailed do we need to be? *Natural Hazards* **49**: 79–98.
- Aricò C, Sinagra M, Begnudelli L, Tucciarelli T. 2011. MAST-2D diffusive model for flood prediction on domains with triangular Delaunay unstructured meshes. *Advances in Water Resources* **34**: 1427–1449.
- Aronica G, Hankin B, Beven KJ. 1998. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources* **22**: 349–365.
- Aronica G, Bates PD, Horritt MS. 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes* **16**: 2001–2016.
- Aronica G, Franza F, Bates PD, Neal JC. 2012. Probabilistic evaluation of flood hazard in urban areas using Monte Carlo simulation. *Hydrological Processes* **26**: 3962–3972.
- Bates PD. 2004. Remote sensing and flood inundation modelling. *Hydrological Processes* **18**: 2593–2597.
- Bates PD. 2012. Integrating remote sensing data with flood inundation models: how far have we got? *Hydrological Processes* **26**: 2515–2521.
- Bates PD, De Roo APJ. 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology* **236**: 54–77.
- Bates PD, Horritt MS, Fewtrell TJ. 2010. A simple inertial formulation of the shallow water equations for efficient two dimensional flood inundation modelling. *Journal of Hydrology* **387**: 33–45.
- Beven KJ. 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* **320**: 18–36.
- Beven KJ, Binley AM. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298.
- Beven KJ, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *Journal of Hydrology* **249**: 11–29.
- Beven KJ, Leedal DT, McCarthy S, Lamb R Hunter M, Keef C, Bates PD, Neal J, Wicks J. 2011a. Framework for assessing uncertainty in fluvial flood risk mapping. *FRMRC Research Report SWP1.7*.
- Beven K, Smith PJ, Wood A. 2011b. On the color and spin of epistemic error (and what we might do about it). *Hydrology and Earth System Sciences* **15**: 2123–2133.

- Blazkova S, Beven K. 2009a. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research* **45**: W00B16.
- Blazkova S, Beven K. 2009b. Uncertainty in flood estimation. *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance* **5**: 325–332.
- Cloke HL, Wetterhall F, He Y, Freer JE, Pappenberger F. 2013. Modelling climatic impacts on floods with ensemble climate projections. *Quarterly Journal of the Royal Meteorological Society* **139**: 282–297.
- Di Baldassarre G, Montanari A. 2009. Uncertainty in river discharge observations: a quantitative analysis. *Hydrology and Earth System Sciences* **13**: 913–921.
- Di Baldassarre G, Schumann G, Bates PD. 2009. A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. *Journal of Hydrology* **367**: 276–282.
- Domeneghetti A, Vorogushyn S, Castellarin A, Merz B, Brath A. 2013. Probabilistic flood hazard mapping: effects of uncertain boundary conditions. *Hydrology and Earth System Sciences* **17**: 3127–3140.
- Dottori F, Todini E. 2013. Testing a simple 2D hydraulic model in an urban flood experiment. *Hydrological Processes* **27**: 1301–1320.
- Dottori F, Di Baldassarre G, Tofini E. 2013. Detailed data is welcome, but with a pinch of salt: accuracy, precision, and uncertainty in flood inundation modelling. *Water Resources Research* **49**: 6079–6085.
- Falter D, Vorogushyn S, Lhomme J, Apel H, Gouldby B, Merz B. 2013. Hydraulic model evaluation for large-scale flood risk assessments. *Hydrological Processes* **27**: 1331–1340.
- Fewtrell TJ, Bates PD, Horritt M, Hunter NM. 2008. Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes* **22**: 5107–5118.
- Fewtrell TJ, Duncan A, Sampson CC, Neal JC, Bates PD. 2011a. Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data. *Physics and Chemistry of the Earth, Parts A/B/C* **36**: 281–291.
- Fewtrell TJ, Neal JC, Bates PD, Harrison PJ. 2011b. Geometric and structural river channel complexity and the prediction of urban inundation. *Hydrological Processes* **25**: 3173–3186.
- Freer JE, McMillan H, McDonnell JJ, Beven KJ. 2004. Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology* **291**: 254–277.
- Gouldby B, Sayers P, Mulet-Marti J, Hassan MAAM, Benwell D. 2008. A methodology for regional-scale flood risk assessment. *Proceedings of the Ice—Water Management* **161**: 169–182.
- Hall J, Tarantola S, Bates P, Horritt M. 2005. Distributed sensitivity analysis of flood inundation model calibration. *Journal of Hydraulic Engineering* **131**(2): 117–126.
- Hall J, Boyce S, Wang Y, Dawson R, Tarantola S, Saltelli A. 2009. Sensitivity analysis for hydraulic models. *Journal of Hydraulic Engineering* **135**(11): 959–969.
- Hall J, W, Manning LJ, Hankin RKS. 2011. Bayesian calibration of a flood inundation model using spatial data. *Water Resources Research* **47**: W05529.
- Horritt MS. 2006. A methodology for the validation of uncertain flood inundation models. *Journal of Hydrology* **326**: 153–165.
- Horritt MS, Bates PD. 2001a. Effects of spatial resolution on a raster based model of flood flow. *Journal of Hydrology* **253**: 239–249.
- Horritt MS, Bates PD. 2001b. Predicting floodplain inundation: raster-based modelling versus the finite-element approach. *Hydrological Processes* **15**: 825–842.
- Hunter NM, Horritt MS, Bates PD, Wilson MD, Werner MGF. 2005. An adaptive time step solution for raster-based storage cell modelling of floodplain inundation. *Advances in Water Resources* **28**: 975–991.
- Hunter NM, Bates PD, Horritt MS, Wilson MD. 2006. Improved simulation of flood flows using storage cell models. *Proceedings of the ICE—Water Management* **159**: 9–18.
- Hunter NM, Bates PD, Neelz S, Pender G, Villanueva I, Wright NG, Liang D, Falconer RA, Lin B, Waller S, Crossley AJ, Mason DC. 2008. Benchmarking 2D hydraulic models for urban flooding. *Proceedings of the ICE—Water Management* **161**: 13–30.
- IPCC. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Stocker TF, Qin D, Plattner G-K, M Tignor, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA; 1535.
- Jamieson SR, Lhomme J, Wright G, Gouldby B. 2012. A highly efficient 2D flood model with sub-element topography. *Proceedings of the ICE—Water Management* **165**: 581–595.
- Jung HC, Jasinski M, Kim J, Shum C, Bates P, Neal J, Lee H, Alsdorf D. 2012. Calibration of two-dimensional floodplain modeling in the central Atchafalaya Basin Floodway System using SAR interferometry. *Water Resources Research* **48**: W07511.
- L'homme J, Sayers P, Gouldby B, Samuels P, Wills M, Mulet-Marti J. (2008). Recent development and application of a rapid flood spreading method. In *Flood Risk Management—Research and Practice*, Samuels P, Huntington S, Allsop W, Harrop J (eds). Taylor & Francis Group: London, UK; 15–24.
- Luke A, Kaplan B, Neal J, Lant J, Sanders B, Bates P, Alsdorf D. 2015. Hydraulic modeling of the 2011 New Madrid Floodway activation: a case study on floodway activation controls. *Natural Hazards* **77**: 1863–1887.
- Mason DC, Bates PD, Dall' Amico JT. 2009. Calibration of uncertain flood inundation models using remotely sensed water levels. *Journal of Hydrology* **368**: 224–236.
- McMillan H, Krueger T, Freer J. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes* **26**: 4078–4111.
- Merz B, Thielen AH. 2009. Flood risk curves and uncertainty bounds. *Natural Hazards* **51**: 437–458.
- Mignot E, Paquier A, Haider S. 2006. Modeling floods in a dense urban area using 2D shallow water equations. *Journal of Hydrology* **327**: 186–199.
- Munich RE (2014). Loss events worldwide 2013. [online]. Available from: http://www.munichre.com/site/corporate/get/documents_E833834344/mr/assetpool.shared/Documents/0_Corporate%20Website/6_Media%20Relations/Press%20Releases/2014/natural-catastrophes-2013-wold-map_en.pdf [Accessed 1 May 2015]
- Neal JC, Bates PD, Fewtrell TJ, Hunter NM, Wilson MD, Horritt MS. 2009. Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. *Journal of Hydrology* **368**: 42–55.
- Neal JC, Schumann G, Fewtrell TJ, Budimir M, Bates PD, Mason D. 2011. Evaluating a new LISFLOOD-FP formulation with data from the summer 2007 floods in Tewkesbury, UK. *Journal of Flood Risk Management* **4**: 88–95.
- Neal J, Villanueva I, Wright N, Willis T, Fewtrell T, Bates P. 2012. How much physical complexity is needed to model flood inundation?. *Hydrological Processes* **26**: 2264–2282. DOI: 10.1002/hyp.8339
- Néelz S, Hall J, Pender G. 2007. Improving the performance of fast flood inundation models by incorporating results from very high resolution simulations. In *Proceedings of flood risk assessment II conference, Institute of Mathematics & its Applications, Southend on Sea* (pp. 1–9).
- Pall P, Aina T, Stone DA, Stott PA, Nozawa T, Hilberts AGJ, Lohmann D, Allen MR. 2011. Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* **470**: 382–385.
- Pappenberger F, Beven KJ, Ratto M, Matgen P. 2008. Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources* **31**: 1–14.
- Parkes BL, Cloke HL, Pappenberger F, Neal J, Demeritt D. 2013. Reducing inconsistencies in point observations of maximum flood inundation level. *Earth Interactions* **17**: 1–27.
- Purvis MJ, Bates PD, Hayes CM. 2008. A probabilistic methodology to estimate future coastal flood risk due to sea level rise. *Coastal Engineering* **55**: 1062–1073.
- Quinn N, Bates PD, Siddall M. 2013. The contribution to future flood risk in the Severn estuary from extreme sea level rise due to ice sheet mass loss. *Journal of Geophysical Research, Oceans* **118**: 5887–5898.
- Romanowicz R, Beven K. 2003. Estimation of flood inundation probabilities as conditioned on event inundation maps. *Water Resources Research* **39**(3): 1073.
- Sampson CM, Fewtrell TJ, Duncan A, Shaad K, Horritt MS, Bates PD. 2012. Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. *Advances in Water Resources* **41**: 1–17.

- Sanders BF. 2008. Integration of a shallow water model with a local time step. *Journal of Hydraulic Research* **46**: 466–475.
- Savage J, Bates P, Freer J, Neal J, Aronica G. 2014. The impact of scale on probabilistic flood inundation maps using a 2D hydraulic model with uncertain boundary conditions. *Vulnerability, Uncertainty, and Risk* **279–289**.
- Siebert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research* **38**(11): 1241. DOI:10.1029/2001WR000978.
- Smith A, Freer J, Bates P, Sampson CM. 2014. Comparing ensemble projections of flooding against flood estimation by continuous simulation. *Journal of Hydrology* **511**: 205–219.
- Stephens E, Bates P. 2015. Assessing the reliability of probabilistic flood inundation model predictions. *Hydrological Processes* **29**: 4264–4283. DOI: 10.1002/hyp.10451.
- Stephens EM, Bates PD, Freer JE, Mason DC. 2012. The impact of uncertainty in satellite data on the assessment of flood inundation models. *Journal of Hydrology* **414–415**: 162–173.
- Stephens E, Schumann G, Bates PD. 2014. Problems with binary pattern measures for flood estimation. *Hydrological Processes* **28**: 4928–4937.
- Tarpanelli A, Brocca L, Melone F, Moramarco T. 2013. Hydraulic modelling calibration in small rivers by using coarse resolution synthetic aperture radar imagery. *Hydrological Processes* **27**: 1321–1330.
- Werner M, Blazkova S, Petr J. 2005. Spatially distributed observations in constraining inundation modeling uncertainties. *Hydrological Processes* **19**: 3081–3096.
- Yu D, Lane SN. 2006. Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 1: mesh resolution effects. *Hydrological Processes* **20**: 1541–1565.